

Association Rules Analysis using FP-Growth Algorithm to Make Product Recommendations for Customer

Ni Putu Priyastini Dessy Safitri

Department of Information Technology,
Faculty of Engineering Udayana University, Bukit Jimbaran, Bali, Indonesia

ABSTRACT

Companies usually have historical data on sales transactions from month to month, but unfortunately, they are only used as weekly and monthly reports. If it is allowed to continue for longer, there will be data growth which results in data richness but poor information. At the same time, companies often still use manual methods in their product marketing strategies that have no reference and are only based on estimates. One of them is the X Fashion Store that sells various local fashions. X Fashion Store has not used data to develop their marketing strategy. This study conducted an association rules analysis to develop a sales strategy. Sales transaction data used is data for December 2020 with a minimum value of support of 25% and a minimum value of confidence of 80% by processing data using Rapidminer application. FP-Growth algorithm can produce association rules as a reference in product promotion and decision support in providing product recommendations to consumers based on predetermined minimum support and confidence values. The association rule result with the highest lift ratio is 10.51.

KEYWORDS: Association, FP-Growth, Product Recommendation

How to cite this paper: Ni Putu Priyastini Dessy Safitri "Association Rules Analysis using FP-Growth Algorithm to Make Product Recommendations for Customer"

Published in
International Journal
of Trend in Scientific
Research and
Development (ijtsrd),
ISSN: 2456-6470,
Volume-5 | Issue-2,
February 2021,
pp.528-531,
www.ijtsrd.com/papers/ijtsrd38459.pdf



Copyright © 2021 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



A. INTRODUCTION

Information technology has been adopted in various sectors, especially in companies, but there are still companies that have not utilized it optimally. Companies usually have historical data on sales transactions from month to month, but unfortunately, they are only used as weekly and monthly reports. If it is allowed to continue for longer, there will be data growth which results in data richness but poor information. This untreated data will only result in a buildup of useless data[1].

Companies often still use manual methods in their product marketing strategies that have no reference and are only based on estimates. This results in the promotion are not on target and promotion costs are wasted. The company's marketing strategy needs to take advantage of computing technology. Stored sales transaction data can provide benefits for company management, for example, to increase sales through an appropriate product sales promotion strategy[2][3]. Various marketing activities, such as promotions and packaging for discounts, can be done better and on target by utilizing this historical data. The majority of consumers are interested in promotional programs because they feel they can save money. Conditions like this can be a means to increase sales, retain customers, and provide satisfaction to customers.

One of the important information from historical data that can be used for the benefit of marketing activities is frequent patterns (FP). This FP contains pattern information that

often appears from transaction data. One technique for extracting FP information is to use the association rule as a data mining method that can identify similarities between items. The Frequent Pattern-Growth (FP-Growth) algorithm can be used to form candidate item combinations to apply associative rules that have a certain frequency value. This algorithm has been implemented in a variety of applications [2], [3]. The importance of the association rule can be seen from two parameters, namely minimum support (the percentage of item combinations of all transactions) and minimum confidence (the strength of the relationship between items in the associative rule).

This FP-Growth algorithm research is different from previous research, this study aims to examine the implementation of the algorithm and association rule to the purchase transaction dataset to provide promotional product recommendations for customers. The implementation was carried out at the X Fashion Store which was established in early 2018 and sells various local fashions. The X Fashion Store has not used data to develop their marketing strategy. The dataset is in the form of sales transaction data for December 2020. The analysis of the highest frequency patterns produces a combination of items that meet the specified minimum support requirements. The formation of the association rule pattern is done to find the association rule that meets the minimum confidence of the high frequencies that have been found.

B. Research Method

Association rules uses a case study in the X Fashion Store. Association rules analysis is done using FP-Growth Algorithm. There are several stages of research for association rules analysis can be seen in Figure 1.

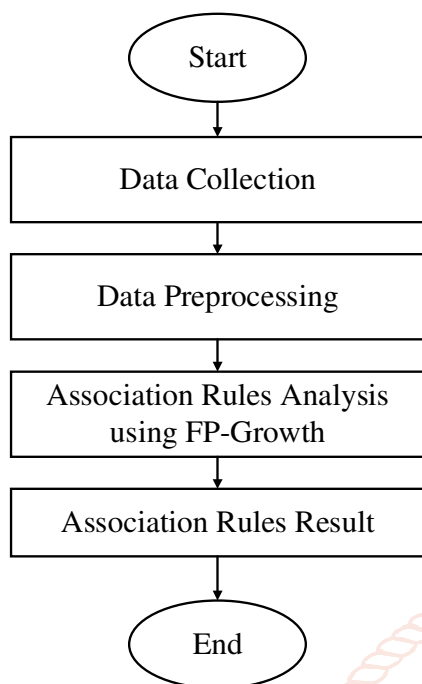


Figure 1 Research Stage

1. Data Collection

The data collection stage is the stage for collecting sales data and literature studies. X Fashion Store does not process data that is owned and has no reference in promoting products. Sales data is used for product recommendations, namely sales transaction data of 246 transactions in December 2020.

2. Data Preprocessing

The preprocessing stage prepares the data obtained from the data collection stage so that the format is by the research needs. Sales transaction data is carried out by a data selection and data transformation. The amount of data previously contained 246 data, after the data selection, it became 184 data. The result is a data representation in the form of a binary matrix data vectorization. This data vectorization matrix was created to adjust the data representation format so that it fits the FP-Growth algorithm data input format in the Rapidminer application. Figure 2 is the example of binary matrix data vectorization.

Open in Turbo Prep Auto Model				
Row No.	Base Hat	Be Dress	Be Jeans	Bing Shirt
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	1	0	0	0
5	0	0	0	0
6	0	0	1	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0

Figure 2 the result of data preprocessing

3. Association Rules Analysis using FP-Growth

The association process is carried out using the Rapidminer application to get the results of the association rules. Rapidminer is software that contains an integrated environment for processing data, machine learning, deep learning, and predictive analysis[2]. The FP-Growth algorithm flow consists of several processes, namely searching for support item values, building FP-Tree, searching for frequent item set, searching for support and confidence values on item set, and obtaining the results of association rules[4]. The flowchart of the FP-Growth algorithm can be seen in Figure 3.

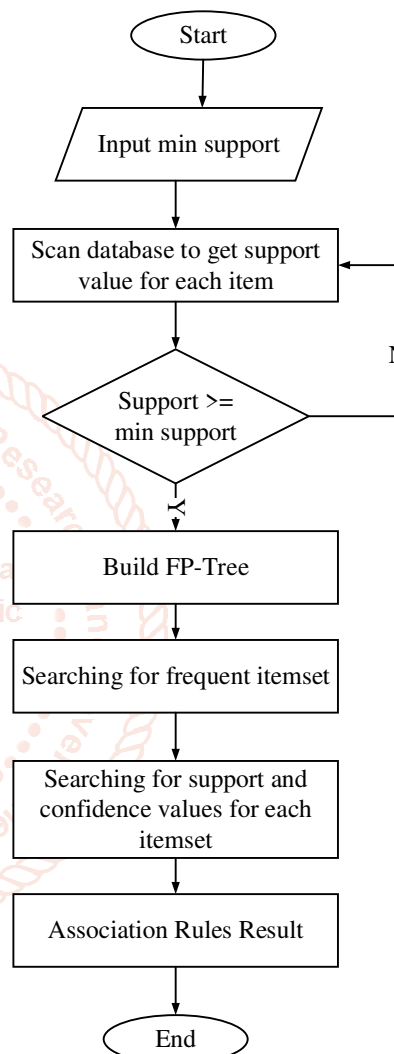


Figure 3 Flowchart of the FP-Growth Algorithm

Association rule is a rule that states the correlation between the occurrence rate of several attributes in a dataset [5]. The terms antecedent and consequent are found in the association rules. Antecedent to represent the "if" part and consequent to represent the "then" part. An example of the form of association rule " $A \rightarrow B$ ", where A is the antecedent and B is the consequent. Antecedent is also called LHS (Left Hand Side) while the consequent is called RHS (Right Hand Side), with the interpretation that every purchase of an item on LHS allows a purchase on RHS[6]. The accuracy of association rules is support, confidence, and lift ratio. Support is the frequency of relationships between items in the data [7]. The support value of an item is obtained by Formula 1.

$$\text{Support}(A) = \frac{\sum \text{Transaction Containing } A}{\sum \text{Transaction Data}} \quad (1)$$

Formula 1 is the formula for finding the support value (A), which is the value of supporting the combination of item A in the database. The number of transactions containing A is the appearance of item A in the entire transaction. The number of transactions is the total number of transactions in the database. The formula used to find the support value of two or more items can be seen in Formula 2.

$$\text{Support}(A, B) = \frac{\sum \text{Transaction containing } A \& B}{\sum \text{Transaksi}} \quad (2)$$

Formula 2 is a formula to find support values (A, B), which is the value of supporting a combination of two items, namely item A and item B in the database. The number of transactions containing A and B is the appearance of items A and B in the entire transaction. Total transactions are the total number of transactions in the database. Confidence is a measurement that shows the degree of certainty or truth of the relationship between items [12]. For example, how often is item B purchased if the consumer buys item A. The confidence value of the association rule $A \Rightarrow B$ is obtained by formula 3.

$$\text{Confidence}(A \Rightarrow B) = \frac{\text{Support } A \Rightarrow B}{\text{Support } A} \quad (3)$$

Lift ratio is a measurement that shows the strong relationship between items in the association rules. The lift ratio formula can be seen in Formula 4.

$$\text{Lift}(A \Rightarrow B) = \frac{\text{Confidence}(A, B)}{\text{Benchmark Confidence}(A, B)} \quad (4)$$

4. Association Rules Result

The results of association rules will use to make Customer Product Recommendations.

C. Literature Study

Previous research conducted by Indrawan and Saputra to Manage Cross-Selling and Up-Selling for IT Shop. The research was conducted using the FP-Growth algorithm to produce association rules with the highest support and confidence, namely Crimping Set and RJ45 Connector with a support value of 0.550 and a Confidence value of 0.785[2].

Research conducted by Ikhwan, Nofriansyah, and Sriani, namely conducting research to support the Education Promotion Strategy by implementing the fp-growth algorithm. Information relating to the implementation of promotions can be quickly available, for example in choosing a place to be used as education promotion, so that the management of the campus itself can make decisions quickly. The resulting association rules are if the student is from a private high school, his address is Medan and the origin of his social studies course, then he chooses the Information Systems Study Program with a 100% confidence level and is supported by 6% of the overall data[3].

Research on the implementation and comparison of algorithms for market basket analysis was carried out by Hidayat. This study uses data from the Breiliant Shop to analyze consumer spending patterns by comparing two algorithms, namely the a priori algorithm and FP-Growth. The result of this research is that the two algorithms can be used to obtain the correlation of product purchase items[5].

D. Result

Association process result in Rapidminer application version 9.8 can be done by entering binary matrix data vectorization. Then run the model as shown in the figure 4.

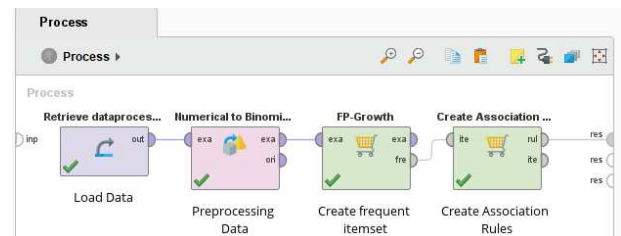


Figure 4 Process Association

From the process of implementing the FP-Growth algorithm with the Rapidminer application above it produces five association rules as follows in the Table 1.

Table 1 Association Rules Results

No	Rules	Supp	Conf	Lift
1	{LightMask, LavJacket} → {KinayJeans}	0,07	0,93	10,51
2	{Je Pants} → {Liss Shirt}	0,08	0,94	8,63
3	{LavJacket} → {KinayJeans}	0,09	0,94	10,22
4	{KinayJeans} → {LavJacket}	0,09	1	10,22
5	{LightMask, KinayJeans} → {LavJacket}	0,07	1	10,22

Table 1 is association rules result such as five of the best association rules with a minimum support of 0,25 and a minimum confidence of 0,8. Rule 1, namely {Light Mask, Lav Jacket} → {Kinay Jeans} shows that of all customers who bought Light Mask and Lav Jacket, 93% also bought Kinay Jeans. Rule 2, namely {Je Pants} → {Liss Shirt} shows that of all customers who purchase Je Pants, 94% also buy Liss Shirts. Rule 3, namely {Lav Jacket} → {Kinay Jeans} shows that of all customers who bought Lav Jacket products, 93% also bought Kinay Jeans. Rule 4, namely {Kinay Jeans} → {Lav Jacket} shows that of all customers who bought Kinay Jeans products, 93% also bought Lav Jacket. Rule 5, namely {Light Mask, Kinay Jeans} → {Lav Jacket} shows that of all customers who bought Light Mask and Kinay Jeans products, 93% also bought {Lav Jacket}.

The higher the lift ratio value in the association rule, the greater the accuracy strength of the association process. The itemset combination in the association rule is valid and strong if the lift ratio value is more than 1 [8].

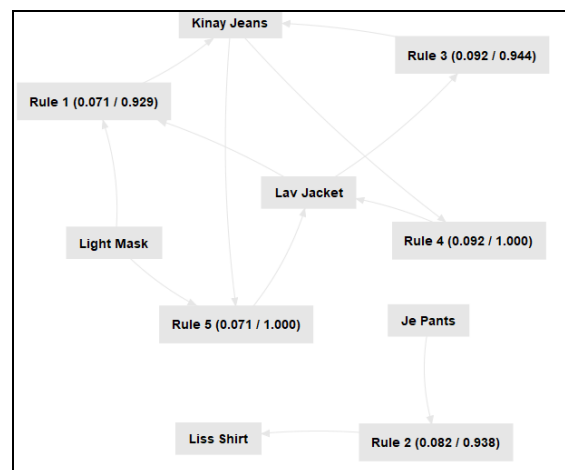


Figure 5 Visualization Association Rules

Figure 5 is visualization of the rules. The visualization image shows that Light Mask, Lav Jacket and Kinay Jeans have a relationship between items that are often purchased by

consumers and there is also a relationship between Je Pants and Liss Shirt items.

The rules formed above can be used as a reference in product recommendations. Marketing strategies can be carried out, for example, product bundling (Light Mask, Lav Jacket and Kinay Jeans will be sold as one saving package), buy 2 get 1 (if customer buy 2 pcs of Je Pants customer will get 1 pcs Liss Shirt), and product discounts (get a discount 5% if you buy Je Pants over 5 pcs). Thus, the company has a reference for handling marketing strategies, not just estimates. The results of this research can be implemented by the company to support the marketing strategy. Efforts to improve results can also be improved, for example by modifying the minimum support and confidence values to get a better one or applying other association algorithms, such as the a priori algorithm.

E. Conclusion

The FP-Growth algorithm can be implemented through stages in sales transaction data. The association rules that are formed can be used as a reference for product recommendations that meet the minimum values of support and confidence. The association rule with the highest lift ratio in the {LightMask, LavJacket} -> {KinayJeans} rule is the lift ratio of 10,51.

References

- [1] R. Ryszky and M. Sadikin, "Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi Pelanggan," *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 3, pp. 103–108, 2019.
- [2] I. W. A. Indrawan and K. O. Saputra, "Implementation of Association Rules to Manage Cross-Selling and Up-Selling for IT Shop," vol. 4, no. 2, pp. 60–63, 2019.
- [3] A. Ikhwan, D. Nofriansyah, and Sriani, "Penerapan Data Mining dengan Algoritma Fp-Growth untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma)," *Saintikom*, vol. 14, no. 3, pp. 211–226, 2015.
- [4] N. Riyadi, M. F. Mulki, and R. Susanto, "Analysis of Customers Purchase Patterns of E-Commerce Transactions Using Apriori Algorithm and Sales Forecasting Analysis With Weighted Moving Average (WMA) Method," *Scientific Research Journal (SCIRJ)*, vol. VII, no. VII, pp. 45–58, 2019.
- [5] A. A. Hidayat, A. Rahman, R. M. Wangi, R. J. Abidin, R. S. Fuadi, and W. Budiawan, "Implementation and Comparison Analysis of Apriori and fp-growth Algorithm Performance to Determine Market Basket Analysis in Breilant Shop," *Journal of Physics: Conference Series*, vol. 1402, no. 7, 2019.
- [6] M. Fauzy, K. R. Saleh W, and I. Asror, "Penerapan Metode Association Rule Menggunakan Algoritma Apriori pada Simulasi Prediksi Hujan Wilayah Kota Bandung," *Jurnal Ilmiah Teknologi Informasi Terapan*, vol. II, no. 2, pp. 221–227, 2016.
- [7] P. N. Tan, M. Steinbach, and V. Kumar, "Association analysis: Basic concepts and algorithms," *Introduction to Data Mining*, pp. 327–414, Jan. 2005.
- [8] D. Fitriati, "Implementasi Data Mining untuk Menentukan Kombinasi Media Promosi Barang Berdasarkan Perilaku Pembelian Pelanggan Menggunakan Algoritma Apriori," in *Prosiding Annual Research Seminar 2016*, 2016, vol. 2, no. 1, pp. 472–480.